

# WORD NETWORKS (PART 1)

A. ROSS ECKLER

Morristown, New Jersey

Most logologists have noticed that some words, such as SHAME and SHARE, are relatively near each other, whereas others, such as TONIC and WATER, are far apart. If pressed for a definition of the distance between two words, they might suggest that it is equal to the number of letters that must be altered to transmute one word into the other. The nearer two words are to each other, the easier it is to accidentally change one into another by a typographical miscue. A large number of English words are only one letter apart, a situation humorously exploited for many years in the Readers Digest column "Pardon, Your Slip Is Showing". Students of the language have suggested that English has more potentiality for mistakes of this sort than other languages do; see, for example, Petr Beckmann's The Structure of Language: A New Approach (Golem Press, 1972) for the development of this idea.

The object of this two-part article is to explore the nearness of words in a systematic fashion. Imagine, for example, all three-letter words written down on a sheet of paper, with lines joining every pair of words that differs in only one letter, such as ERA and BRA, or BAT and BUT, or SEE and SEA. One can ask many questions about the properties of this word network -- for example, what word has the largest number of lines emanating from it? are all words in a single network, or are there several networks isolated from each other, like an archipelago?

The answers to these questions depend, of course, on the number of words one includes in the study. The difficulties of characterizing word networks are in general so great that it has been necessary to severely restrict the vocabulary. In particular, we use only those words that appear in boldface type in the main section (pp. 1-592) of the Merriam-Webster Pocket Dictionary. Hyphenated words, suffixes, prefixes and abbreviations are omitted, as are words listed without further definition (anti-, in-, over-, re-, self-, sub-, super- and un-). Even so, the vocabulary size is so large that it is feasible to show actual networks only for two-letter words and (in part) three-letter words in the first part of this article. The vocabulary sizes for various word-lengths are given below:

2-letter words: 38	4-letter words: 1843	7-letter words: 4591
3-letter words: 541	5-letter words: 2798	8-letter words: 4415
	6-letter words: 4056	

The task of searching for words near other words was greatly facilitated by a new positional word list, Volume 1 of the Word Builder's Handbook, published by Computer Puzzle Library, P.O. Box 5134, Fort Worth, Texas 76108. Volume 1, covering words from 2 to 8 letters in length, sells for \$12 but a special offer of \$9.90 is sometimes available (see March 1973 General Contest Bulletin). Volume 2, priced similarly, covers words from 9 to 15 letters. The claim is made that these volumes were prepared with meticulous care using modern computer capabilities to provide the greatest possible accuracy, and indeed the accuracy is greater than in other publications of this type. Computer processing insures that the various rearranged lists all contain the same words with none accidentally left out, but it does not guard against keypunch errors in preparing the original list. Two such keypunch errors: EUREKA is spelled UEREKA, and YCLEPT is spelled YCELPT.

In the two parts of this article, we study in some detail seven different word networks, consisting of words from two to eight letters in length. Broadly speaking, most of the shorter words of a given length are joined in a single network, whereas the longer words of a given length are scattered among a number of independent networks. It is for this reason that word ladders (chains of the form LESS-LOSS-LOSE-LORE-MORE) are usually restricted to words of five or fewer letters. The chance is quite small that a random pair of seven-letter or eight-letter words can be joined by a word ladder in the Pocket Dictionary.

### Two-Letter Word Network

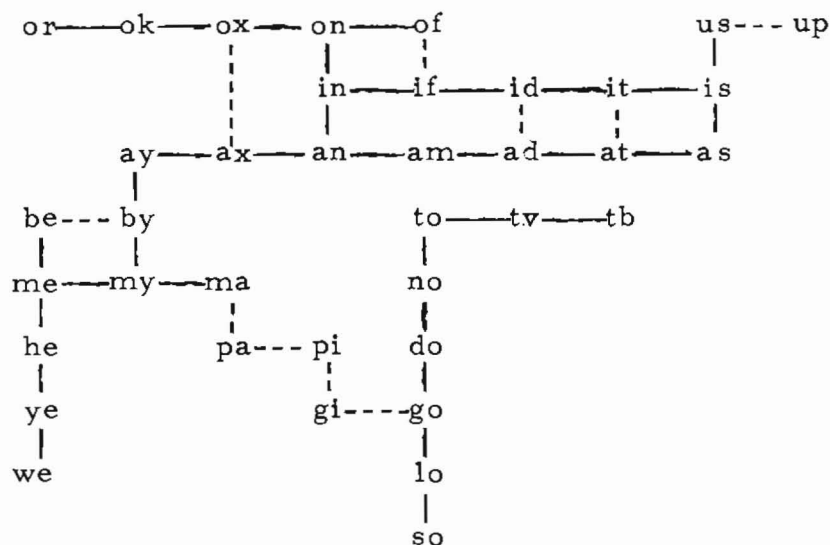
As there are only 38 two-letter words in the Pocket Dictionary, it is an easy matter to map the entire network. In the diagram on the next page, a group of three or more words differing only in their final letter is listed in a row and joined by solid lines; similarly, a group of three or more words differing only in their first letter is listed in a column and joined by solid lines. Solid lines act as a kind of shorthand for the considerably more numerous lines that should be plotted between every pair of words in a group; for example, the fact that one can move from OR to OF by means of parallel solid lines should be interpreted as meaning that one can move from OR to OF in a single step, leapfrogging OK, OX and ON. Groups of only two words are joined by dotted lines.

The most important feature of this network is that no two-letter words are left out. Note, however, that a slightly different vocabulary would have broken it into two independent networks -- all that is necessary is to remove GO, or GI, or PI, or PA, or MA.

It is useful to introduce here several properties of a word network that are helpful in visualizing larger networks that cannot be so easily diagrammed. What word in the network has the maximal ambiguity -- that is, what word can be transmuted into the largest number of other words with the change of only one letter? AN can be transmuted into

the eight words IN, ON, AY, AX, AM, AD, AT and AS; similarly, AS and AY have eight near neighbors.

The densest part of the network can be identified not by a single word but by a group of words in a rectangular array, such as OX-AX-ON-AN. Such patterns have previously been called garble groups in



Word Ways; in response to a query, Leslie Card exhibited various maximum-size garble groups in the November 1970 and February 1971 issues. In the above network, it is easy to locate the largest garble group: IN-AN-ID-AD-IT-AT-IS-AS. Since the first letter can take on either of two values and the second letter can take on any of four values, a shorthand notation for the shape of the garble group is given by (24). Note that the maximum garble group includes two of the three maximally ambiguous words.

A slightly more sophisticated concept is the span of a word network. For any two words in a network, there is a shortest path joining them -- for example, the shortest path between MA and AX is MA-my-ay-AX with three steps, although the five-step path MA-me-be-by-ay-AX is also possible. As the initial and final words are varied, how long can such a shortest path be? Examining the network, it is not hard to see that the shortest path joining UP to TB has eleven steps: UP-us-as-ay-my-ma-pa-pi-gi-go-to-TB. A similar number of steps is required to join UP to TV, OR to TV or TB, OF to TB or TV, and OK to TV or TB. Thus, the span of this word network is eleven.

There are  $(26)(26) = 676$  possible ways that words of two letters can be formed. The word network of 38 words occupies less than six per cent of this space. How widely scattered through the space are the words in the network? One measure of scatter was introduced by Dmitri Borgmann in "Irrelevance" in Beyond Language (Scribner's, 1967). How many words can be found in the

network with the property that all the first letters of the words are different from each other and all the second letters of the words are also different from each other? Put another way, how many words can be found which have a maximum resistance to ambiguity with each other? A little experimentation reveals that there are several different ways that ten such words can be selected; one such set is TV, NO, GI, PA, MY, HE, AX, OK, IN and US.

### Three-Letter Word Networks

The three-letter word network is considerably more complicated than the two-letter one. To begin with, there is no single network which includes all 541 three-letter words; the twelve words listed below are isolated from the main network and from each other:

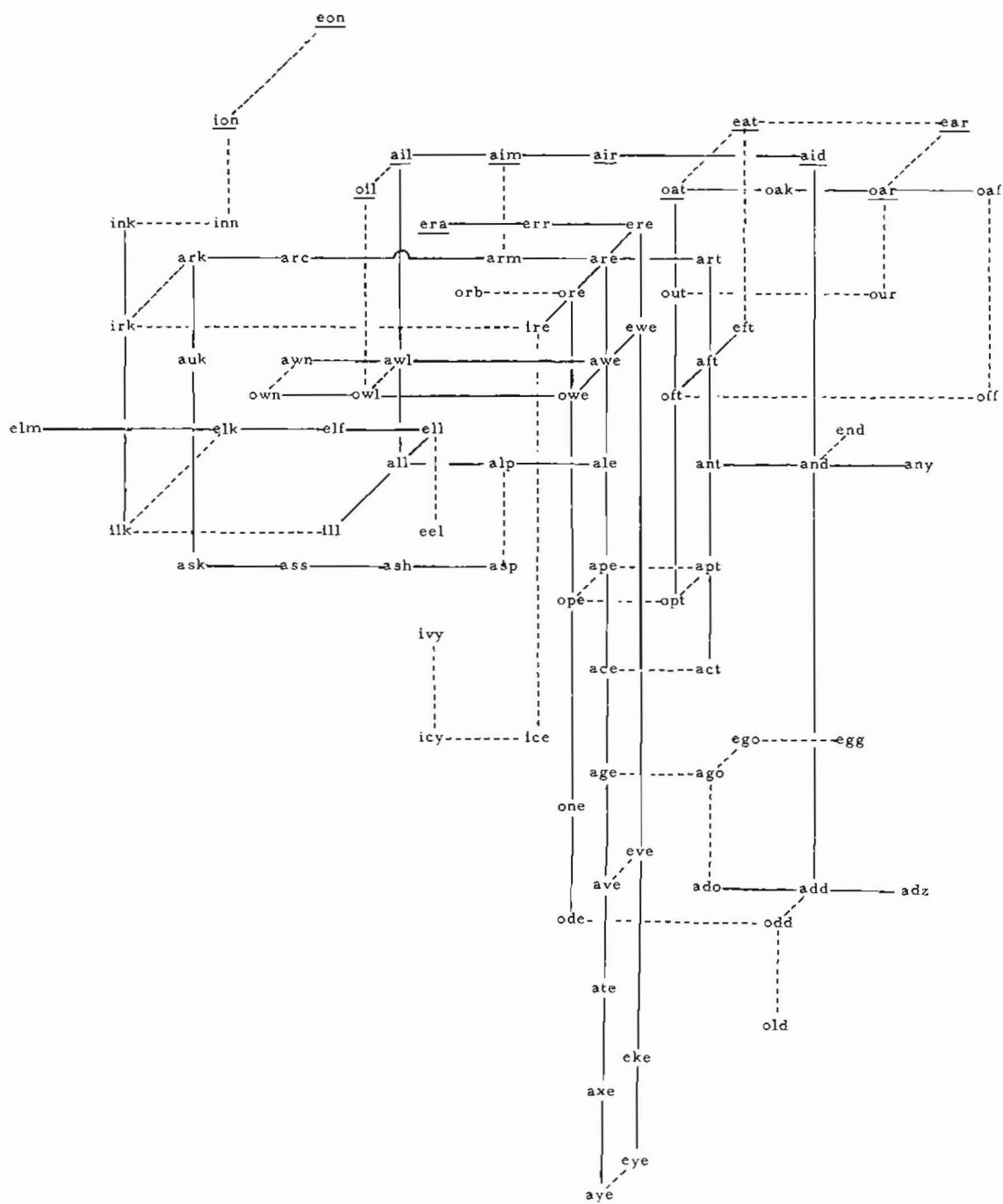
ebb	gnu	ism	nth	ohm	urn
emu	imp	its	obi	ova	use

In the February 1969 Word Ways, Rudolph Castown proposed the term singularity for a word which does not connect with any other word, and the term isolano for a group of words which do not connect with the main word network (i.e., the largest network). Later, Dave Silverman used isolano for a single isolated word, and ignored the problem of naming an isolated group. We shall adopt the latter nomenclature, and note that there are twelve three-letter isolanos in the Pocket Dictionary.

The remaining 529 words from a single network which is far too complicated to diagram. To give some flavor of the network, we have diagrammed on the next page those 86 words which begin with the vowels A, E, I and O (there are no words beginning with U other than the two isolanos above), the sparsest section of the network. Note that the network is now three-dimensional -- a group of words differing only in the third letter is listed in a row, a group of words differing only in the second letter, in a column, and a group of words differing only in the first letter, along a diagonal meant to suggest a vertical to the page. It may be helpful to think of the network as the somewhat incomplete steel framework of a building with four floors -- words beginning with I in the basement, then words beginning with O, next words beginning with A, and finally words beginning with E on the roof.

The twelve underlined words at the top of the diagram are the only places where this part of the network connects with words beginning with consonants. Already, this hints at a property that will become more pronounced in later networks -- it is relatively difficult to cross over from words beginning with vowels to words beginning with consonants, and words beginning with vowels are far more likely to form isolanos or isolated groups.

To get some idea of the undiagrammed dense part of the network, let us consider the various properties introduced earlier. The net-



work word having maximal ambiguity is almost certainly SAY with 25 lines joining it: SAC, SAD, SAG, SAP, SAT, SAW, SHY, SKY, SLY, SOY, SPY, STY, BAY, CAY, DAY, FAY, GAY, HAY, JAY, LAY, MAY, NAY, PAY, RAY and WAY. (In the diagrammed part, one can easily check that the most ambiguous word is ARE with 16 connections.)

This network is exceptionally rich in large garble groups. The largest one is believed to be the (523) group given below:

cap	cop	hap	hop	map	mop	sap	sop	tap	top
cat	cot	hat	hot	mat	mot	sat	sot	tat	tot
caw	cow	haw	how	maw	mow	saw	sow	taw	tow

The second and third best garble groups are given below; their shapes are (515) and (432), respectively. Note that the (515) group fails to exploit one of the available dimensions; all of its words have A as their second letter:

cad	cap	cat	caw	cay	hat	pat	sat	tat
had	hap	hat	haw	hay	hit	pit	sit	tit
mad	map	mat	maw	may	hot	pot	sot	tot
pad	pap	pat	paw	pay				
sad	sap	sat	saw	say	hap	pap	sap	tap
					hip	pip	sip	tip
					hop	pop	sop	top

Note that HAP, HAT, SAP and SAT occur in all three groups, suggesting the general area of the densest part of the network. (There are no three-dimensional garble groups in the diagrammed part, although there are numerous two-dimensional groups of sizes 4 and 6, and one of size 8.)

The span of this word network is somewhat difficult to determine. It is conjectured that no word-pairs exist having a greater minimum path-length than IVY and YOU: IVY-icy-ice-ire-are-arm-aim-dim-din-win-won-yon-YOU is one of several equivalent twelve-step paths. Notice that the span is barely greater than that of the two-word network, although the vocabulary size is more than ten times as large. Clearly, it is a much more densely connected network than the earlier one was.

The number of words having maximum resistance to ambiguity rises from ten to seventeen: one such list is ADD, BRA, CAR, DEW, EBB, FLY, GNU, HIT, IMP, JOG, NTH, OWL, PYX, RUN, SKI, USE and WHO. Word lists of this type are apt to make disproportionate use of isolanos; five of the twelve are represented here. The theoretical maximum number of such words is 26, so that this list is quite a respectable one.

# Four-Letter Word Networks

Only 2.2 per cent of all three-letter words in the Pocket Dictionary are isolated from the main network; this percentage rises to 10.5 when four-letter words are considered. The isolanos begin to have an interesting group structure of their own; for example, one group of eleven words is unconnected to the main network. A listing of all isolano groups follows:

adze	apex	doff	huge	kyat	nuts	ovum	upon
agog	apse	drys	hymn	liar	obey	raja	urdu
ague	arch	echo	ibex	lira	oboe	ruby	urge
ahoy	ashy	egis	iced	luau	odds	stye	uric
alga	auld	envy	iffy	lynx	odor	suds	user
alms	aura	epee	inky	magi	okay	sync	veto
amok	avow	epic	iota	memo	okra	tabu	whom
ammo	awol	espy	jehu	meow	once	taxi	whys
anal	bias	evil	judo	myna	onyx	thou	xmas
ankh	crux	gaff	kayo	nazi	orgy	tyro	zebu
	dais	hobo	khan	nova	ours	ugly	

afar-ajar	army-arty	edam-exam	mama-maya
amah-ayah	axes-axis	ibis-iris	ogle-ogre
ante-anti	does-goes	idol-idyl	oleo-olio

anus-onus	anew-knew	also-alto	
opus	knee	auto	icon-ikon-iron

bevy-levy	idly-idle	eddy-edgy	opal-oral-oval
leva	isle	edge	
	dodo-dido		quad-quai-quay
	lido		

aged			
abed-abet	into-onto-unto	ache-acme-acne-acre	
abut	undo		

each-etch	nigh	ally	
itch-inch	high	ably	
inca	sikh-sigh	abbe-able	
huff	sign	axle	
ruff			
luff tiff	edit	oxen	
muff-miff	exit	open	
puff	omit-emit	eden omen-amen	
cuff	obit	emir	even-oven
buff			ever-over-aver
			ewer

The isolation of words beginning with vowels is quite marked. All but 46 of the 171 vowel words are in the isolano groups above, and the rest are connected to the main network only in small clusters.



All clusters involving more than two words are diagrammed below.

WIRY	area-urea	
airy	GRID-arid-aria	
ZERO-aero-aery-eery-VERY	acid	
awry	amid	VASE-ease-else
away	avid	easy
SWAY		VAST-east-erst

STEM-item-idem-idea-ides

However, that part of the network consisting of words beginning with consonants appears to be fairly strongly connected. Most of these words are of the form CVCC, CVVC or CVCV (where C stands for a consonant and V a vowel); there are many bridges between each pair of these patterns.

The network word having maximal ambiguity is probably WARE, which has 24 lines joining it: BARE, CARE, DARE, FARE, HARE, MARE, PARE, RARE, TARE, WERE, WIRE, WORE, WADE, WAGE, WAKE, WALE, WANE, WAVE, WARD, WARM, WARN, WARP, WART and WARY.

Four-letter garble groups are markedly smaller than three-letter ones, suggesting that the densest part of the four-letter network is somewhat more open than the densest part of the three-letter network. Two sixteen-word garble groups are known, with shapes of (4221) and (4141):

dane	fane	mane	wane	bale	bane	bare	bate
dine	fine	mine	wine	dale	dane	dare	date
				male	mane	mare	mate
dare	fare	mare	ware	pale	pane	pare	pate
dire	fire	mire	wire				

Note that neither of these garble groups has a pattern in which there are changes of all letters. In geometric terms, there is no garble group from the Pocket Dictionary that exploits all four dimensions. Although garble groups with the shape (2222) are possible (see Castown's article on word ladders in the August 1968 issue), words not in the Pocket Dictionary must be used. The nearest approach using Pocket Dictionary words appears to be:

bard	band	bare	bane	ward	wand	ware	wane
----	bond	bore	bone	word	----	wore	----

The number of words having maximum resistance to ambiguity remains at the high level of seventeen: AMMO, CZAR, ETCH, FIZZ, HUSK, IKON, KNEW, LYNX, NEWT, OGLE, PLUM, RAJA, TWIG, URDU, WHYS, SPRY and YOGI. One should note that only eight of these words come from the main network.